Towards an intuitive human-robot interaction based on hand gesture recognition and proximity sensors

Gorkem Anil Al, Pedro Estrela and Uriel Martinez-Hernandez

***Abstract*— In this paper, we present a multimodal sensor interface that is capable of recognizing hand gestures for human-robot interaction. The proposed system is composed of an array of proximity and gesture sensors, which have been mounted on a 3D printed bracelet. The gesture sensors are employed for data collection from four hand gesture movements (*up*, *down*, *left* and *right*) performed by the human at a predefined distance from the sensorised bracelet. The hand gesture movements are classified using Artificial Neural Networks. The proposed approach is validated with experiments in offline and real-time modes performed systematically. First, in offline mode, the accuracy for recognition of the four hand gesture movements achieved a mean of 97.86%. Second, the trained model was used for classification in real-time and achieved a mean recognition accuracy of 97.7%. The output from the recognised hand gesture in real-time mode was used to control the movement of a Universal Robot (UR3) arm in the CoppeliaSim simulation environment. Overall, the results from the experiments show that using multimodal sensors, together with computational intelligence methods, have the potential for the development of intuitive and safe human-robot interaction.**

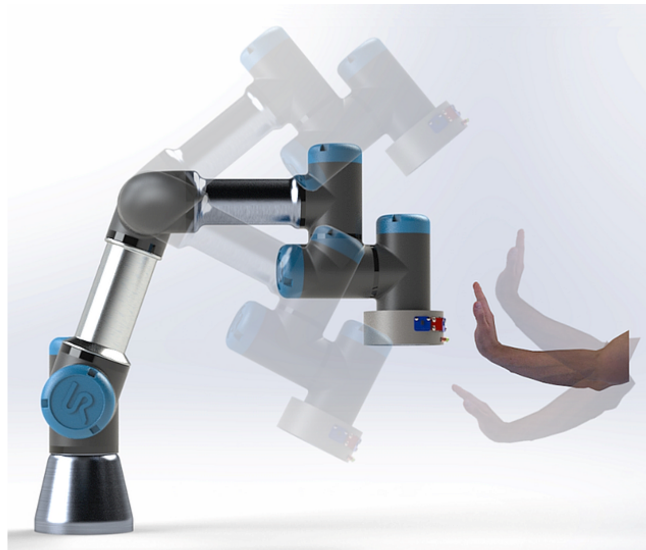


Fig. 1. Multimodal sensor array mounted on a 3D printed bracelet for controlling the movements of the UR robot based on the recognition of hand gestures performed by the human.

I. INTRODUCTION

Robotics is expected to play a crucial role for the transition from traditional to flexible manufacturing, where robots are capable of moving autonomously, safely interacting with humans and easily programmable by non-experts to perform multiple tasks [1]. In human-robot collaborative tasks, the design of intuitive interfaces is crucial to benefit the robot from the human experience and skills to perform complex tasks. These interfaces should be intuitive and easy to use for humans to interact with robots without any concern, while making them feel comfortable to work on their tasks [2].

For efficient and safe human-robot collaboration, the end-users (operators) need to be able to control the robots in an intuitive way, while the robot accurately perceives the human actions and commands. These processes require the use of multiple sensing modalities and computational intelligence methods, and different strategies have been studied in previous works for human-robot interaction and collaboration, e.g., voice, touch, vision and user interfaces [3], [4], [5].

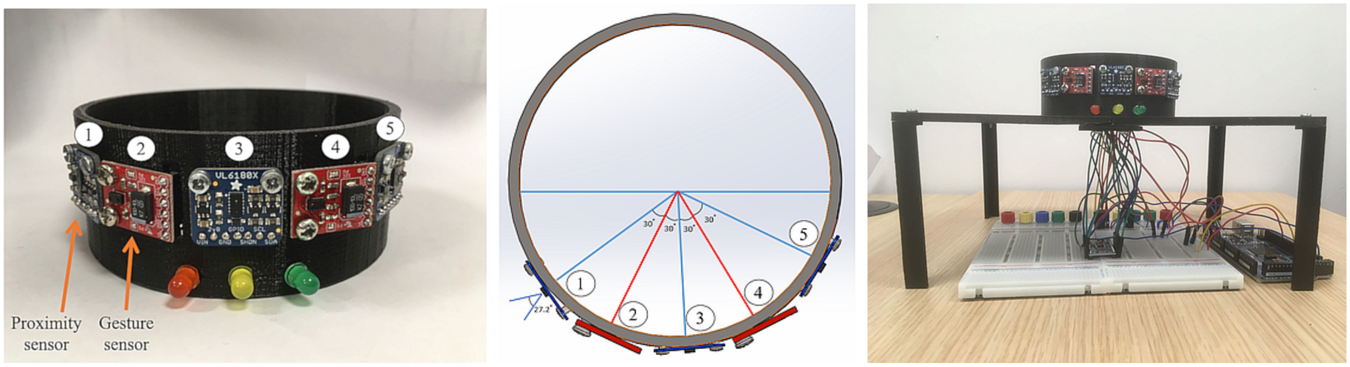
Voice recognition has been used to command industrial robots for pick-and-place and welding operations [6]. Although verbal communication channels are useful for human-robot interaction, this method is not appropriate for industrial environments given the noise levels that affect the recognition of voice commands [2].

Vision-based techniques with depth cameras have been used for the study of the interaction between humans and robots based on tracking the human operator body and hand gestures [7], [8], [9]. Stereo cameras have also been employed for the recognition of hand gestures such as number and pointing [5], [10]. These works are constrained to fixed places due to the required set of cameras. The combination of voice and gesture sensing has been investigated to allow the robot to have a better understanding of the human movements, and thus, control the robot position [11], [12].

Non-vision and wearable systems have been used to identify sets of body movements for robot control in human-robot interaction [13]. Static and dynamic human hand behaviours, captured with the Nintendo Wii Remote controller, have been employed for robot programming [14]. However, the Wii controller can be cumbersome and interfere with the hands of users in industrial applications. Haptic technology, with tactile and force sensors, has been used to control the compliance of robots while interacting with humans in home and industrial environments [15], [16], [17]. This approach offers dynamic interaction features with robots compared with vision-based gesture interaction, which requires different setup based on changes in the work environment.

*This work was supported by The Republic of Turkey Ministry of National Education, and the Royal Society Research Grants for the 'Touching and feeling the immersive world' project (RGS/R2/192346).

[1]Gorkem, Pedro and Uriel are with Centre for Autonomous Robotics (CENTAUR), and Centre for Bioelectronics, Biosensors and Biodevices (C3Bio), Department of Electronic and Electrical Engineering, University of Bath, UK (gga31, p.estrela, u.martinez)@bath.ac.uk



(A) (B) (C)

Fig. 2. Multimodal sensorised bracelet. (A) Time of Flight and gesture sensors labelled as 1, 3, 5 and 2 and 4, respectively. LEDs mounted for visual inspection of the data collection and recognition processes. (B) The sensors are mounted at a distance of 30° from each other. (C) Experimental setup composed of the sensorised bracelet and 3D printed table designed for systematic data collection and hand gesture recognition experiments.

In this work, a multimodal sensor interface for safe human-robot interaction using proximity and gesture sensors is presented. This approach uses a contactless hand gesture recognition technology that have not been applied in the literature to provide accurate and safe human-robot interaction. The multimodal sensor interface is responsible for the recognition of hand gestures (*up*, *down*, *left*, *right*) in real-time and control the movement of a robot platform (Figure 1). This approach allows the human to interact and control the robot platform without touching it or physically contacting it while performing a task in a shared workspace. The multimodal sensors are mounted on a 3D printed bracelet designed for the Universal Robot (UR3). A computational method based on an Artificial Neural Network (ANN) is used for the recognition of human hand gestures. The validation of this work is performed with experiments in off-line and real-time mode. In off-line mode, the recognition accuracy of hand gestures is tested with ANNs and multimodal sensor data. In real-time mode, hand gestures are recognised using multimodal sensor data in real-time to control the UR3 robot in a simulation environment. Overall, these experiments demonstrate the capability of multimodal interfaces, together with computational intelligence, for safe and intuitive robot control in human-robot interaction tasks.

II. Methods

A. Design of the multimodal sensorised bracelet

The multimodal sensor interface for human-robot interaction is composed of a 3D printed bracelet and an array ToF and gesture sensors. The 3D printed bracelet has been designed to be mounted on the end-effector of the UR3 robot (Figure 1). The ToF sensor VL6180x, composed of IR emitter, range and ambient modules, is employed to detect the proximity of the human to the robot platform between 150 mm and 250 mm. The sensor is capable of measuring up to 600 mm distance and 290 mm diameter with 27.2° viewing cone. The gesture sensor APDS-9960 is used for data collection from the hand movements performed by the human located in front of the bracelet for interaction with the robot. The APDS-9960 sensor also provides proximity information, however, it has lower sensitivity than the VL6180x sensor. The APDS-9960 sensor uses four photodiodes to collect data from hand movements. The reflected IR energy, sourced by integrated LEDs, is converted from motion to digital information. A set of red, yellow and green LEDs are mounted on the 3D printed bracelet for testing purposes to visually indicate the state of the robot, whether the human is at the correct distance for data collection and whether the data collected is ready for the recognition process. This multimodal sensor interface, arrangement and position of the sensors are shown in Figures 2A and 2B.

The multimodal sensor data is transmitted to a workstation for data analysis using an Arduino microcontroller, which communicates to the sensors via I^2C and using the multiplexer TCA9548 from Texas Instruments. This multiplexer board includes eight channels and can be connected to a programming card using one I^2C bus. The experimental setup composed of the multimodal 3D printed bracelet and the circuitry for communication is shown in Figure 2C.

B. Data acquisition

The data collection task was performed by one user using the experimental setup showed in Figure 2C. The set of hand gestures employed for data collection is composed of *up*, *down*, *left* and *right* hand movements. The raw data from the hand gestures performed by the human was obtained from the four photodiodes internally located in the APDS-9960 sensor. The values of this data are in the range of 0 and 255 according to the distance from the human hand to the sensor. The length of raw data collected from the sensors depends on the speed of the hand gesture movement. The data collected from the four hand gesture movements has been normalised and examples of the data collected are shown in Figure 3.

Gesture data was collected by moving the hand in front of the sensorised bracelet in a distance between 150 mm and 190 mm. This distance was defined according to the gesture sensor operating range which is between 100 mm and 200 mm. This was used to define the appropriate range of distances, between 150 mm and 190 mm, for data collection.

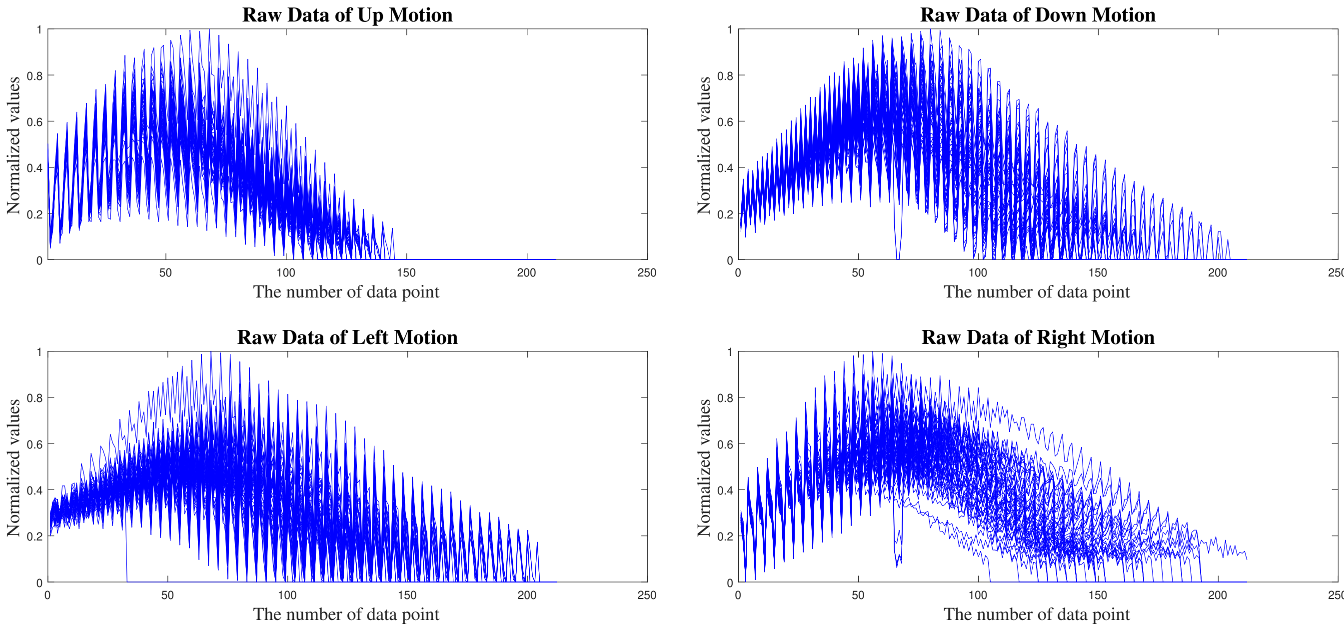


Fig. 3. Example of raw data collected from one gesture sensor from four hand gestures movements.

The set of steps for data collection from a hand gesture is as follows. First, the hand moves towards the bracelet, when the hand is in the range of 200 mm to 250 mm the yellow LED turns on to visually indicate that the hand is moving toward the bracelet. Second, when the hand is close to the bracelet in the range of 150 mm to 190 mm, the green LED turns on to indicate that the system is ready and that a hand gesture (*up*, *down*, *left* or *right*) can be performed to start the data collection process. The red LED is used only during the hand gesture recognition process to inform the operator whether the recognition has been successful or not. The data collection process was repeated 100 times for each hand gesture, generating datasets to train and test the model for the recognition of hand gestures. In the current approach, the 3D printed bracelet uses two gesture sensors to collect data for the recognition process. Additionally, 100 data samples obtained from one gesture sensor only were labelled as error trials.

C. Artificial Neural Network for hand gesture recognition

Computational intelligence has demonstrated its potential for recognition processes in a variety of robotic applications [18], [19], [20]. In this work, a supervised learning approach using an Artificial Neural Network (ANN) is employed for the recognition of hand gesture movements. This ANN receives as input the raw data obtained from the two gesture sensors embedded in the 3D printed bracelet. Then, the ANN provides as output one of the four hand movements performed by the human. The length of the raw data obtained from the sensorised bracelet changes from a minimum and maximum number of data points of 26 and 496 per sensor, respectively, depending on the speed of the hand movement. Therefore, multiple lengths of the raw data, with increments of 20 data points, were used to train the ANN and identify the optimal data length for the recognition process. Thus, the ANN was trained with input datasets of dimensions of 500×52 (100 samples per gesture and 52 data points from 2 sensors) to 500×992 with increments of 20 data points.

Three different ANNs, composed of one hidden layer with 5, 10 and 15 neurons, were implemented to analyse the accuracy for the recognition process. These ANNs were implemented with the conjugate gradient optimisation algorithm for the training phase. The input data was segmented into 70% for training, 15% for validation and 15% for testing each ANN. The ANNs were trained and tested five times with each dataset in order to calculate the mean recognition accuracy over the five iterations. The training and testing phases of the ANNs for the recognition of hand gestures with multimodal data were implemented in MATLAB.

III. Results

The proposed multimodal sensor interface for human-robot interaction has been validated with experiments in offline mode using MATLAB, and real-time mode using real sensor data and the UR robot in a simulation environment.

A. Offline recognition of hand gestures

The experiments in offline mode were performed to analyse the accuracy for recognition of hand gestures using the real data from the sensorised bracelet with different configuration of ANNs in MATLAB. Figure 4 shows the flowchart with the processing steps to receive the sensor data and prepare it for the ANN. The raw data of the two gesture sensors from Arduino are read as char format, and converted to number format. In MATLAB, the raw data is divided into sensorv1 and sensorv2 vectors that contain the data from gesture sensor 1 and gesture sensor 2, respectively. The length of raw data is built with 26 samples from each sensor.

If the length of the collected data from any gesture sensor is less than 26 samples, then the data is padded with zero values. If the length of raw data is larger than 26 samples, then only the first 26 data samples are used. This process ensures that the training and testing dataset have the correct length. Next, the data from both sensors is concatenated to form the input vector of 1×52 dimension for the ANN.

The mean recognition accuracy over the four hand gestures (*up*, *down*, *left*, *right*) was evaluated using ANNs with 5, 10 and 15 neurons in the hidden layer. The results in Table I show that the ANN with 5 hidden neurons and 500×52 input data size achieved the highest mean recognition accuracy of 97.86%. The lowest accuracy of 96% was achieved with the ANN composed of 15 hidden neurons and an input data of dimensions 500×132. The performance from the training, validation and testing phases over the four hand gestures with the ANN with 5 neurons, and input data of 500×52 are shown in Figure 5A. The accuracy for recognition of individual hand gestures is presented with the confusion matrix in Figure 5B. These results in offline mode validate the potential of multimodal sensor arrays for the recognition of hand gestures, which can be used for robot control.

B. Real-time recognition of hand gestures for robot control

The recognition of hand gestures in real-time mode was performed using real-time data, the ANN with 5 hidden neurons and 25 repetitions of each hand gesture direction (*up*, *down*, *left* and *right*). The ANN with 5 hidden neurons was selected based on the highest accuracy achieved in the offline mode experiment. The results from the recognition of individual hand gestures in real-time mode are presented in the confusion matrix shown in Figure 5C. These results show that the *right* and *down* hand gestures were recognised with a mean accuracy of 100%, while the *up* and *left* hand gestures achieved mean accuracies of 95.7%, and 94.7%, respectively.

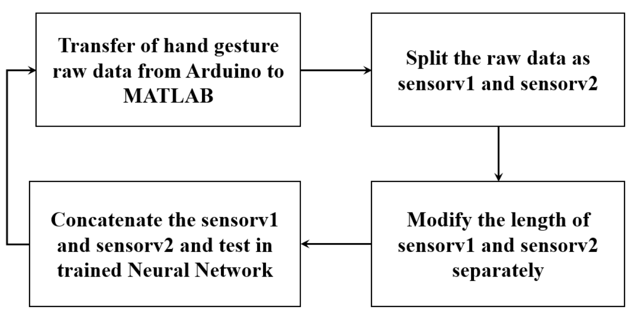


Fig. 4. Diagram with the steps performed to transfer the sensor data from Arduino to MATLAB, and to prepare the data for recognition by the ANN.

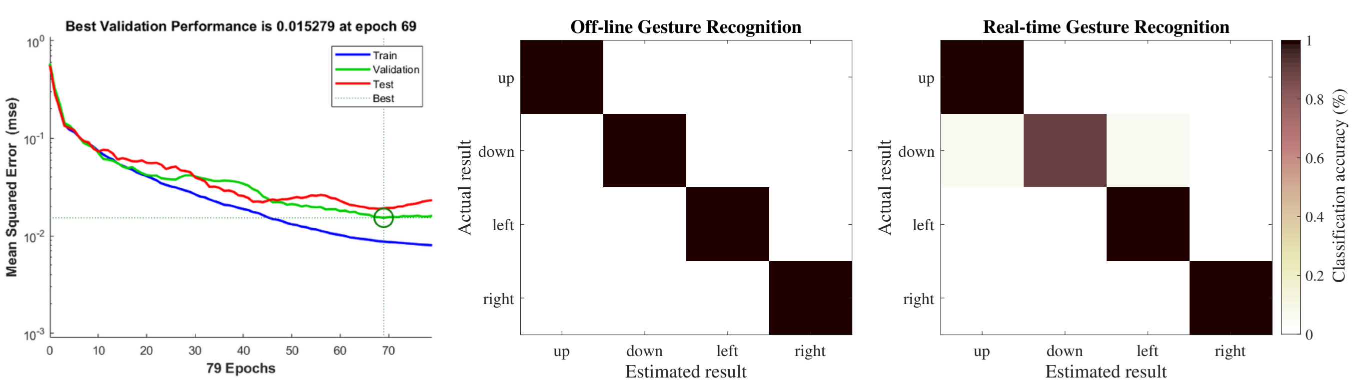
TABLE I

MEAN ACCURACY RECOGNITION OF HAND GESTURES WITH MULTIMODAL SENSOR DATA AND THREE DIFFERENT ANNS

	5 Neurons	10 Neurons	15 Neurons
Data Size	500x52	500x372	500x132
Confusion Matrix Accuracy	97.86%	97.06%	96%

The real-time experiment was also performed in a scenario where the human commands a robot arm to move using hand gestures during an assembly task in a simulation environment (see Figure 6). This experiment permitted to evaluate the potential of the proposed approach for human-robot interaction in industrial environments. For this experiment, real data was collected from the multimodal sensor interface and the recognition output from the ANN was used to control the UR3 robot arm in the CoppeliaSim simulation environment. Each of the four hand gestures *up*, *down*, *left* and *right* were mapped to the centre of the orange, blue, green and yellow areas, respectively, shown in Figure 6A. For example, when the operator needs a tool from the orange area, then, the operator can command the robot to move to the orange area by performing the *up* hand gesture movement. The same process is followed for the other hand gestures.

Safety should be guaranteed given that the robot and the operator work in the same environment. Therefore, the recognition process and the robot respond to the hand gesture when the human operator is in a distance to the robot between 150 mm and 190 mm. When the distance between robot and the operator is in the range of 200 mm and 250 mm, the robot stops and waits for the next command from the operator. The panel of LEDs mounted on the sensorised bracelet have been included for testing purposes



(A) (B) (C)

Fig. 5. Results from the hand gesture recognition experiments in offline mode. (A) Training, validation and testing performance with the ANN implemented with 5 neurons in the hidden layer. (B) Confusion matrix with the recognition accuracy of each individual hand gesture. (C) Confusion matrix with the accuracy for recognition results of individual hand gestures performed in real-time mode by the ANN.

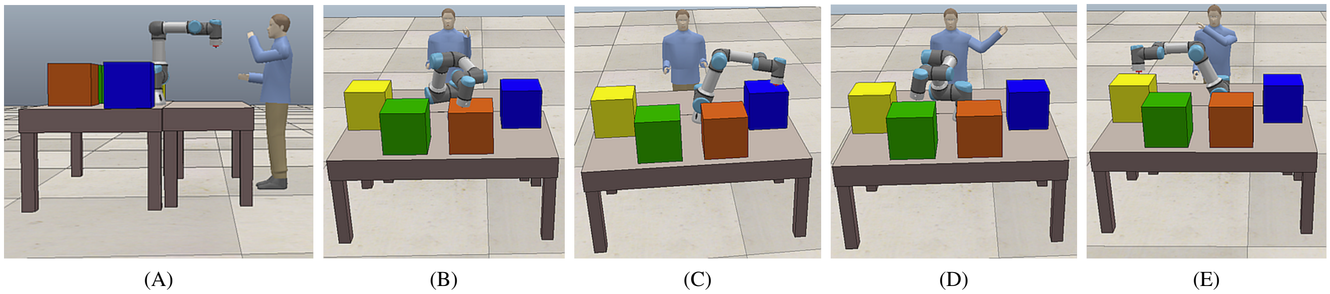


Fig. 6. Human and robot interaction based on hand gestures in simulation environment. (A) The robot at home position waiting for hand gesture. (B-D) The robot moves to the orange, blue, orange green and yellow areas when the human performs the *up*, *down*, *left* and *right* hand gesture, respectively.

and indicate the state of the interaction process. This is another safety feature that allows the human operator to visually know whether the robot arm is ready to perform another action. The yellow LED turns on when the distance between the operator and the robot is between 200 mm and 250 mm. When the operator moves his/her hand closer to the sensorised bracelet in a distance between 150 mm and 190 mm, the green LED turns on, indicating that the system is ready for data collection from the hand gesture. The red LED turns on when the hand gesture performed by the human has been recognised correctly, then the output is send to the simulation environment to control the movement of the UR3 robot arm. This interaction process between the human and robot is as follows:

- The robot is in home position with the sensorised bracelet facing towards the human and waiting for the hand gesture command
- The robot recognises the command and executes the action to move towards the corresponding area of the working environment (e.g., take a tool for the operator)
- The robot moves back to home position (e.g., give the tool to the human) and waits for the next hand gesture command from the operator

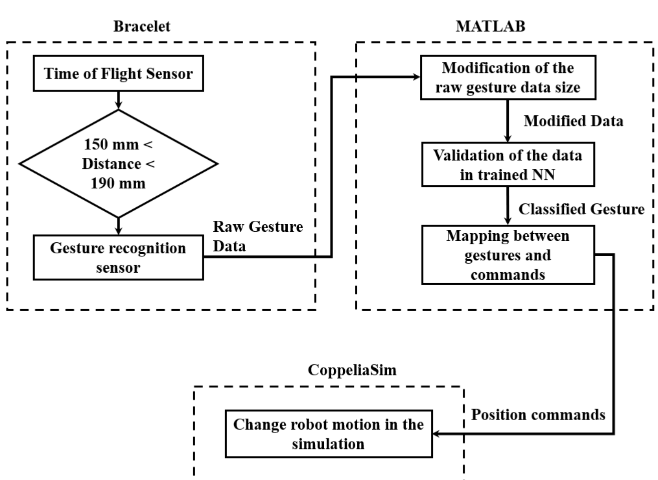


Fig. 7. Framework of the real-time experiments. Bracelet layer; sensors used to read the data from hand gestures. MATLAB layer; data processing for recognition of hand gestures using ANNs in MATLAB. CoppeliaSim layer; UR3 robot controlled by hand gestures in the simulation environment.

The flowchart in Figure 7 shows the processes implemented for the human-robot interaction scenario. The raw data from the gesture sensors are sent from Arduino to MATLAB via serial communication. In MATLAB, the size of the data is processed for the ANN. The joint angles of the robot were set to movement towards each of the coloured areas, according to each hand gesture, using forward kinematics.

Table II shows the comparison between the proposed multimodal sensor interface and different technologies that detect human motion for human-robot interaction. The camera technology can recognise a large set of gesture, however, this approach is limited by occlusion, fixed setup and high computational cost aspects. Voice commands offer a safe interaction, but the recognition process is affected by the noise in industrial environments. Haptic/tactile devices offer an alternative for robot control, but some of these devices require the operator to hold the controllers at all times. Attaching the presented interface on a robot can offer interaction capability from different positions without limiting the motion of the operator, flexibility and less effort to set up the robot when defining new tasks in different work environment. Also, contactless touch can be beneficial for the users when robots are guided to the specific tasks and positions where humans and robots perform collaboration scenarios sequentially.

Overall, the experiments in offline and real-time modes, together with a robotic platform in a simulation environment, have shown the potential of the proposed method for multimodal recognition of hand gestures with applications to human-robot interaction tasks.

IV. Conclusions

This work presented a method for human-robot interaction using an array of gesture and time of flight sensors to control a robot, based on a set of hand gestures performed by a human operator. This array of sensors was mounted on a 3D printed bracelet designed for the Universal Robot (UR3) as an interface for interaction with humans. The proposed method was tested using four hand gesture movements (*up*, *down*, *left* or *right*), which were accurately recognised using an Artificial Neural Network. Multiple experiments were performed in offline and real-time mode for validation of the multimodal hand recognition approach for human-robot interaction. The experiments in offline mode were capable

TABLE II

COMPARISON OF DIFFERENT HRI TECHNOLOGIES

HRI Technology	Advantages	Limitations
Voice Recognition [6]	Does not restrict operator's mobility, safe contactless interaction	Limited voice diffusion in noisy manufacturing environment
Stereo camera [5], [10]	Capability of detecting different range of gestures, safe contactless interaction	Prone to occlusion, restricted detection area, affected negatively from poor lighting condition, fixed setup
Depth sensing camera [8], [9]	Capability detecting different range of gestures, safe contactless interaction	Prone to occlusion, restricted detection area, affected negatively from poor lighting condition, fixed setup
Remote controller [14]	No occlusion, no dependency on lighting and environmental factors	Restricts mobility, time-consuming setup
Haptic/tactile devices [15], [16], [17]	Dynamic setup on the robot, cost effective, no occlusion	Complex computation to recognize contact gesture, complex robot control needed
Proposed multimodal sensor approach (Gesture-proximity sensors)	Computationally effective, safe for dynamic human motion, safe contactless interaction, dynamic setup on robot, does not restrict the human motion	Gesture recognition can be affected by changes in the light conditions

of achieving a mean hand gesture recognition accuracy of 97.86%. This analysis was performed using ANNs with different number of neurons in the hidden layer, and the configuration that achieved the highest accuracy was used for the experiment in real-time mode. For the experiment in real-time mode, sensor data from the four hand gestures was collected and recognised with a mean accuracy of 97.7%. Then, the output from the recognition process in real-time mode was employed to control the movements of a robotic platform in a simulation environment. This validation was performed to evaluate the potential of the proposed multimodal sensor interface for human-robot interaction using hand gestures in an industrial environment.

Overall, the proposed initial multimodal sensor interface, together with computational intelligence methods, showed the potential for the accurate recognition of hand gestures that can be employed for the development of safe and intuitive for human-robot interaction tasks. In the future work, different sensing modalities will be added, and the interface design will be improved to allow the interaction between the human and robot from longer distances. Beside this, complex hand gestures will be identified on the new design. Additionally, this future work will also be tested on real robot applications.

REFERENCES

[1] M. R. Pedersen, L. Nalpantidis, R. S. Andersen, C. Schou, S. Bøgh, V. Krüger, and O. Madsen, "Robot skills for manufacturing: From concept to industrial deployment," *Robotics and Computer-Integrated Manufacturing*, vol. 37, pp. 282–291, 2016.

[2] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, pp. 248–266, 2018.

[3] M. Petit, S. Lallée, J.-D. Boucher, G. Pointeau, P. Cheminade, D. Ognibene, E. Chinellato, U. Pattacini, I. Gori, U. Martinez-Hernandez, *et al.*, "The coordinating role of language in real-time multimodal learning of cooperative tasks," *IEEE Transactions on Autonomous Mental Development*, vol. 5, no. 1, pp. 3–17, 2012.

[4] S. Mokaram, J. M. Aitken, U. Martinez-Hernandez, I. Eimontaite, D. Cameron, J. Rolph, I. Gwilt, O. McAree, and J. Law, "A ros-integrated api for the kuka lbr iiwa collaborative robot," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 15 859–15 864, 2017.

[5] K. Nickel and R. Stiefelhagen, "Visual recognition of pointing gestures for human–robot interaction," *Image and vision computing*, vol. 25, no. 12, pp. 1875–1884, 2007.

[6] G. Bugmann and J. N. Pires, "Robot-by-voice: experiments on commanding an industrial robot using the human voice," *Industrial Robot: An International Journal*, 2005.

[7] R. Weitschat, J. Ehrensperger, M. Maier, and H. Aschemann, "Safe and efficient human-robot collaboration part i: Estimation of human arm motions," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1993–1999.

[8] P. Barattini, C. Morand, and N. M. Robertson, "A proposed gesture set for the control of industrial collaborative robots," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2012, pp. 132–137.

[9] M. R. Pedersen, D. L. Herzog, and V. Krüger, "Intuitive skill-level programming of industrial handling tasks on a mobile manipulator," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 4523–4530.

[10] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, "A hidden markov model-based continuous gesture recognition system for hand motion trajectory," in *2008 19th International Conference on Pattern Recognition*. IEEE, 2008, pp. 1–4.

[11] I. Maurtua, I. Fernández, A. Tellaeche, J. Kildal, L. Susperregi, A. Ibarguren, and B. Sierra, "Natural multimodal communication for human–robot collaboration," *International Journal of Advanced Robotic Systems*, vol. 14, no. 4, p. 1729881417716043, 2017.

[12] P. A. S. Srimal, M. V. J. Muthugala, and A. B. P. Jayasekara, "Deictic gesture enhanced fuzzy spatial relation grounding in natural language," in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2017, pp. 1–8.

[13] S. Fani, S. Ciotti, M. G. Catalano, G. Grioli, A. Tognetti, G. Valenza, A. Ajoudani, and M. Bianchi, "Simplifying telerobotics: wearability and teleimpedance improves human-robot interactions in teleoperation," *IEEE Robotics & Automation Magazine*, vol. 25, no. 1, pp. 77–88, 2018.

[14] P. Neto, J. Norberto Pires, and A. Paulo Moreira, "High-level programming and control for industrial robotics: using a hand-held accelerometer-based input device for gesture and posture recognition," *Industrial Robot: An International Journal*, vol. 37, no. 2, pp. 137–147, 2010.

[15] A. Muxfeldt, J. N. Haus, J. Cheng, and D. Kubus, "Exploring tactile surface sensors as a gesture input device for intuitive robot programming," in *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, 2016, pp. 1–4.

[16] J. Saenz and M. Fritsche, "Tactile sensors for safety and interaction with the mobile manipulator valeri," in *Proceedings of ISR 2016: 47st International Symposium on Robotics*. VDE, 2016, pp. 1–7.

[17] D. Hughes, J. Lammie, and N. Correll, "A robotic skin for collision avoidance and affective touch recognition," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1386–1393, 2018.

[18] Q. Gao, J. Liu, Z. Ju, Y. Li, T. Zhang, and L. Zhang, "Static hand gesture recognition with parallel cnns for space human-robot interaction," in *International Conference on Intelligent Robotics and Applications*. Springer, 2017, pp. 462–473.

[19] U. Martinez-Hernandez, A. Damianou, D. Camilleri, L. W. Boorman, N. Lawrence, and T. J. Prescott, "An integrated probabilistic framework for robot perception, learning and memory," in *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2016, pp. 1796–1801.

[20] P. Rani, C. Liu, N. Sarkar, and E. Vanman, "An empirical study of machine learning techniques for affect recognition in human–robot interaction," *Pattern Analysis and Applications*, vol. 9, no. 1, pp. 58–69, 2006.